achung

A Probabilistic Model of Oil Discovery

 $\ddot{\phantom{a}}$ 

 $\ddot{\phantom{0}}$ 

 $\sim$ 

James L. Smith

March 1979 Revised: January, 1980 Energy Laboratory Working Paper No. MIT-EL 80-005WP

ę. bl *Bti* $\sim$  $\zeta$ 

## A PROBABILISTIC MODEL OF OIL DISCOVERY

James L. Smith

March, 1979 Revised: January, 1980

The author is indebted to Paul Eckbo for his extensive contributions to this research; and to Hendrik Houthakker, Frank O'Carroll for helpful comments on an earlier draft. Excellent research assistance was provided by Geoffrey Ward. Of course, any errors are the sole responsibility of the author.

This research was made possible by financial support from the Norwegian Research Council for Science and Humanities, the Center for Petroeconomic Studies of the Ch. Michelsens Institute, and the Center for Energy Policy Research of the Massachusetts Institute of Technology. All opinions, findings, and conclusions expressed herein are those of the author, and do not necessarily reflect the views of these organizations.

 $\frac{a_{\rm c}}{2}$  $\hat{\mathcal{F}}$ 

### A PROBABILISTIC MODEL OF OIL DISCOVERY

## 1. Introduction

Diverse approaches to the problem of petroleum supply forecasting have been proposed since the appearance of Fisher's pioneering study [8]. Proposed techniques include both engineering-based methods of resource assessment [14] and the more familiar techniques of econometric extrapolation [5,6,7,11,12], and some combinations of the two [13,16]. Regardless of method, however, all studies have experienced the common difficulty of accounting satisfactorily for the phenomenon of exploration and discovery, which is crucial in the determination of future supplies. Structural disparities among the alternative models of resource discovery have been described elsewhere [4], and competing forecasts have been shown to be highly sensitive to model specification and the period of estimation [18], and to the degree of regional disaggregation of the data base [7]. The resulting discrepancies pose serious problems for the forecaster, because there is no strong a priori basis for choosing among the alternative specifications.

Recently, Kaufman and his associates have developed a new method of discovery modeling that provides a simple and appealing structural description of the process of resource discovery and depletion  $[1,2,10]$ . Unlike previous methods, Kaufman's formulation provides a lucid and internally consistent theoretical framework against which the reasonableness of resulting forecasts may be judged. However, although the empirical usefulness of the method has been demonstrated by a recent

application to the North Sea petroleum basin [4], its implementation relies heavily on highly specialized techniques of numerical analysis which make it difficult and costly to apply on a broad scale.

The present paper presents a discovery model that has been adapted from that of Kaufman with the purpose of simplifying its empirical application. The formal changes, described below in Section 2, are motivated by two considerations: (1) they significantly reduce the computational demands placed on the forecaster, and (2) they reduce the sensitivity of the resulting estimates to the type of measurement error that is inherent in available reservoir data. The resulting model is applied in Sections 3 and 4 to the North Sea petroleum province, and estimates of remaining reserves and future discoveries are obtained and compared to other estimates taken from the trade press. Section 5 concludes with a brief summary of results. The data used in the North Sea application are presented in an appendix.

## 2. The Discovery Model

The discovery model outlined below constitutes a stochastic production function which governs the relationship and timing between exploratory effort and reservoir discovery. The relationship is stochastic because it allows that a given level of exploratory effort may or may not result in successful discoveries, or perhaps in discoveries of varying size. The relationship is also dynamic, in that the productivity of exploratory effort evolves through time according to physical laws of resource depletion.

The following statistical postulates, first suggested by Kaufman

 $-2-$ 

and his associates, describe the evolutionary nature of the discovery phenomenon:

- (1) The discovery of reservoirs in a petroleum play<sup>1</sup> can be modeled statistically as sampling without replacement from the underlying population of reservoirs.
- (2) The discovery of a particular reservoir from among the existing population is random, with probability of discovery being proportional to reservoir size.

Kaufman employs the two postulates in conjunction with a presumed lognormal size distribution of underlying reservoirs. However, there is nothing in the postulates themselves which confines analysis to the lognormal case. In its place, we assume here that the original deposition of reservoirs follows an arbitrary but discrete size distribution. In general, the resource base may consist of J possible reservoir sizes, denoted  $(S_1, ..., S_J)$ , occurring with respective frequencies  $(n_1, ..., n_J)$ .<sup>2</sup>

The probability that the first discovery  $(D_1)$  will be of size S<sub>1</sub> may then be computed directly from the second postulate:

(1) 
$$
P(D_1 = S_j) = \frac{n_j \cdot S_j}{J}
$$
; for  $j = 1, ..., J$ .  
\n
$$
\sum_{k=1}^{J} n_k \cdot S_k
$$

 $\hat{\alpha}$ 

As exploration and subsequent depletion proceed, the first postulate dictates the way in which the discovery probabilities evolve. In

 $<sup>1</sup>A$  "play" is defined to be a group of similar geological configurations</sup> generated by a series of common geological events, forming a population of structures that is conceived or proven to contain hydrocarbons.

 $2$ In empirical applications, the size of each reservoir is generally measured by the volume of reserves that can be recovered using current technology. We follow this convention in the North Sea application discussed below.

general, the number of reservoirs of size S, remaining prior to the i $\cdot^{\text{th}}$ discovery is determined by the depletion that has gone before. We represent the cumulative number of discoveries of the  $j<sup>th</sup>$  size which precede the i<sup>th</sup> discovery by the symbol  $m_{i,i}$ . The probability that the i<sup>th</sup> discovery will be of the j<sup>th</sup> size, conditional on the sizes of preceding discoveries, may then be written:

(2) 
$$
P(D_i = S_j | D_1, ..., D_{i-1}) = \frac{(n_j - m_{ij}) \cdot S_j}{\sum_{k=1}^{n} (n_k - m_{ik}) \cdot S_k}
$$
; for  $j = 1, ..., J$ .

To compute the discovery probabilities at each stage in the discovery sequence (and the joint probability of the entire sequence) it is necessary to know both the original deposition  $\{n_i\}$  and the number of preceding discoveries of each size  ${m_{i,j}}$ . The latter are determined immediately upon specification of the discovery sequence in question. For convenience, we denote the size index of the i<sup>th</sup> discovery by the symbol  $I_i$ . An arbitrary sequence of N discoveries may then be represented by  $(I_1, ..., I_N)$ .

Conditional upon knowledge of the original deposition, it is then possible to compute the likelihood of any particular discovery sequence by taking the product of successive conditional probabilities:

(3) 
$$
L(I_1, ..., I_N | n_1, ..., n_J) = \prod_{i=1}^N \frac{(n_{I_i} - m_{i1}) \cdot S_i}{\sum_{\substack{\Sigma \\ j=1}} (n_j - m_{i1}) \cdot S_j}
$$

Alternatively, given an observed discovery sequence,  $(I_1, ..., I_N)$ , the likelihood function can be evaluated at alternative points in the space of  $(n_1, ..., n_J)$ , and the original deposition  $(n_1, ..., n_J)$  identified which maximizes the likelihood of having observed the discovery sequence in question. The resulting estimates  $\{n_i\}$  comprise conventional maximum likelihood estimates of the resource base, and bear the usual optimality properties. Such estimates have been computed for the North Sea petroleum basin, based on the sequence of the first 99 known discoveries. These estimates are presented below in Section 3.

Having obtained estimates of the original deposition based on the first N discoveries, it is possible by repeated application of Equation 3 to generate a sequence of predictive probability distributions which characterize the volume of each ensuing discovery. However, this procedure becomes prohibitively expensive as the sequence is extended into the future. Alternatively, the predictive probability distributions may be approximated by simulating the exploratory process according to the two specified discovery postulates, and observing the relative frequency of alternative outcomes. One simulated discovery sequence of this type, based on the North Sea estimates, is presented in the next section.

## 3. North Sea Results

The forecasting technique described above is used here to generate estimates of the resource base and the sequence of future discoveries in the North Sea petroleum play. In the present context, the area of the play is defined geographically as that section of the North Sea lying between  $56^{\circ}$  and  $62^{\circ}$  north latitude and east of the Shetland Islands. Consequently, the resulting North Sea reserve estimates are

-5-

comparable in geographical extent to those cited elsewhere  $[3.4, 20]$ .

To implement the estimation procedure a seven-cell reservoir size distribution is specified, as indicated in Table 1. The choice of a specific categorization scheme is necessarily arbitrary, but has been made with the intent to reflect the fairly broad range of commonly occurring North Sea reservoir sizes.<sup>2</sup>



TABLE 1: RESERVOIR SIZE CLASSIFICATION AND ESTIMATES

reservoir size measured in million barrels.

The data which support the estimation consist of the first 99 discoveries declared in the North Sea -- covering the period from

 $1$ In geological terms, the North Sea play logically extends north of the 620 parallel. However, this region has been deliberately withheld from exploration by the Norwegian government. Therefore, the estimates presented here describe the resource potential of a circumscribed area of the play. Presumably, the area north of the 62<sup>0</sup> parallel will eventually provide a discovery sequence of its own which may then be used to make inferences regarding the underlying resources of that area. A geological description of the western and southern boundaries of the play is found in [3].

 $2$ Any categorization of reservoirs will in general distort the aggregate volume of reserves declared to date. The categories shown here overstate established reserves by approximately three percent.

July, 1967 to December, 1976. Both the order in which the discoveries occurred and their respective magnitudes are considered in deriving the historical discovery sequence,  $(I_1, ..., I_{qq})$ . The specific data which enter the sequence are presented in the appendix.

Given the historical discovery sequence, estimates of the original deposition have been obtained which maximize the likelihood of the observed sequence.<sup>2</sup> These estimates appear in Table 1. As one would expect, the majority of reservoirs appear to be of the smallest size, with relatively few of the larger sizes. The skewed size distribution which results follows a traditional skewed pattern of mineral deposition. However, the distribution does not conform well to the familiar lognormal hypothesis. The discrepancy can be quantified by examining the  $\chi^2$  goodness-of-fit statistic. When the observed cell frequencies (from Table 1) are compared to the expected cell frequencies corresponding to a lognormal deposition, the computed  $x^2$  statistic assumes the value 543.38, with seven degrees of freedom.<sup>3</sup> This outcome would ordinarily lead to rejection of the lognormal hypothesis. However, the test must be interpreted with caution because the "observed" cell frequencies have actually been estimated in this case, rather than observed.

-7-

 $1$ The size of each discovery is taken to be the combined declared volume of recoverable oil and gas reserves, expressed in terms of oil equivalent on a BTU basis.

 $2$ The point of maximum likelihood was identified directly by a grid-type search over the discrete parameter space.

 $3$ The particular lognormal distribution chosen for comparison was taken to have mean and variance equal to that of the observed distribution, after applying Sheppard's correction for grouped data.

By subtracting the observed number of historical discoveries from the estimated deposition, the remaining number of reservoirs in each size class may be inferred. The resulting volume of reserves remaining to be discovered is recorded in the last column of Table 1. Perhaps the most striking aspect of these results is that no additional discoveries are projected among the three largest size categories (i.e., greater than 400 million barrels). Evidently, the North Sea play is in an advanced state of development, such that the probability is small that additional large reservoirs have been overlooked. This finding is consistent with that of Robinson and Morgan [20], who forecast only one additional discovery exceeding 450 million barrels in the British sector. Of course, it is entirely possible that additional large reservoirs will be found in areas of the North Sea which have been excluded from the present analysis (e.g., north of the  $62^{\circ}$  parallel or west of the Shetlands).

A useful check on the validity of our model results is to ask whether the hypothesized discovery process, when applied to the estimated deposition of reservoirs, would indeed generate the reserve volume which our data sources report to have been established by the 99 historical discoveries. By simulating the discovery process many times, we have found the expected volume of reserves established by the first 99 discoveries (as implied by the model) to equal approximately 33.3 billion barrels, with standard deviation equal to 0.77 billion barrels. This conforms closely to the reported volume of 32.6 billion barrels actually established by the 99 discoveries. Thus, the model is at least successful in reproducing past history.<sup>1</sup>

-8-

 $1$ The simulation procedure is described below in Section 4.

The volume of ultimately recoverable North Sea petroleum reserves is estimated to be approximately 43 billion barrels of oil and gas equivalent (Table 1). Table 2 places this number in perspective relative to several judgmental estimates of North Sea reserves made by members of the petroleum industry. It is clear that opinions vary widely within the industry, ranging from 35 to 67 billion barrels, with greatest weight attached to the lower end of this interval. If an industry consensus could be established, it would appear that our estimate would not deviate greatly from it. However, all of the estimates are subject to considerable margins for error and should only be viewed as indications of the likely order of magnitude.

TABLE 2: COMPARISON OF RESULTS

Reserve Estimates (billion barrels)



Source: the Mobil estimate is cited in [20, p. 13]; other industry estimates are cited in [17, p. 25].

By comparing the maximized value of the likelihood function with the reductions that result from imposing constraints on the total volume of ultimately recoverable reserves, it is possible to establish 90% fiducial limits corresponding to our point estimate of 43.2 billion barrels. The bounds obtained extend from 38.4 billion to 65.0 billion barrels. It

is interesting but coincidental that this range corresponds so closely to the range of uncertainty reflected in the industry estimates reported in Table 2. The 95% fiducial limits for our point estimate extend from 38.2 billion to 68.5 billion barrels.

#### 4. Future Discoveries

The behavior of future North Sea discoveries has been simulated by repeated sampling from the estimated remaining population of reservoirs shown in Table 1. The simulation procedure consists of sequential sampling of individual reservoirs from among the remaining population according to the two basic discovery postulates. That is, a complete sequence of discoveries is obtained by successive drawings, without replacement, from the remaining reservoirs, where each reservoir has discovery probability proportional to its size. In total, 10,000 complete discovery sequences have been generated using Monte Carlo methods, and the results are summarized in Table 3.

In generating each discovery sequence, one additional reservoir size category has been appended to account for the possibility of dry holes  $(D_i=0)$ . The probability of striking a dry hole is assumed to be constant (75%) at each step in the discovery sequence. This dry hole risk factor is roughly consistent with historical experience in the North Sea play.

Average discovery size at each step in the projected sequence

 $-10-$ 

 $^{\text{1}}$ We obtain the fiducial limits by invoking the approximate  $\chi^2$  distribution of the likelihood ratio statistic. The constrained maximum of the likelihood function corresponding to a specified reserve level (R) was computed by a grid search over the subspace of  $(n_1, ..., n_J)$  satisfying  $\Sigma$ n.  $S_i$  = R. **·** J J J

# TABLE 3: SIMULATED FUTURE NORTH SEA DISCOVERIES, PARTIAL SEQUENCE

# (million barrels oil equivalent)



lincludes unproductive wells (dry-holes)

<sup>2</sup> excludes unproductive wells

 $\bar{\mathbf{z}}$ 

 $\rm^3$  includes discoveries established by all preceding wells in the sequence

(column two) is computed from all 10,000 drawings (wells), including both productive reservoirs and dry holes. The phenomenon of "discovery size decline" is evident as one proceeds down the sequence. The rate of decline is not dramatic, however, since most of the large fields have already been found, leaving a more homogeneous distribution of reservoirs in the ground. The coefficient of variation of discovery size (column three) measures the degree of risk at each step in the sequence. The coefficient of variation is of substantial magnitude throughout the sequence due to the presumed risk of dry holes.

Conditional discovery size, which is the average of only those wells that were actually successful (approximately 2,500 in number), is also reported at each step in the sequence (column four). The figures indicate the expected discovery size for a well at any point in the sequence, conditional on its striking a reservoir. The coefficient of variation is also reported (column five), and is now smaller due to the exclusion of dry holes. Nevertheless, even the risk attending successful efforts is substantial, with standard deviation being of the same order of magnitude as the expected value.

It is apparent from the conditional discovery sizes that the petroleum industry can expect only modest returns from future exploratory activity. Reservoirs as small as 70 million barrels are not far distant from the threshold size that is necessary in the North Sea to support economically viable development and production at current price levels. (The reader is referred to Eckbo, Jacoby, and Smith [4]). An important implication is that the tax policies and incentive structures adopted by

 $-12-$ 

the British and Norwegian governments could have a substantial effect on the exploration for and recovery of these future reserves. Even a relatively small distortion created by the fiscal regimes could easily trigger a far-reaching decision to abandon the search for addition petroleum in this area.

Finally, we report the average and coefficient of variation of cumulative discoveries, computed at each step in the discovery sequence (columns six and seven). For example, the next 25 wells are expected-to establish a total of 565.9 million barrels of oil and gas equivalent (including dry holes), with standard deviation roughly half that magnitude. In general, these figures measure the combined success of the next j wells, and give some insight into the productivity and risk of sustained drilling programs.

#### 5. Concluding Remarks

 $\ddot{\phantom{1}}$ 

A major strength of the probabilistic discovery model is the simple yet consistent theoretical treatment of depletion and its influence on the productivity of exploration through time. Application of the model provides direct estimates of the physical returns to continued exploratory activity. Supplemented by estimates of exploratory costs and other economic parameters, the results may be used to infer the economic returns to exploration, and to identify future levels of exploration, discovery, and production that are likely to be pursued by the oil industry. One application of this type of comprehensive forecasting approach has been described elsewhere [4].

The present formulation of the discovery model retains Kaufman's

-13-

two basic discovery postulates, but reparameterizes the target population in a manner that facilitates empirical applications. Of course, the simplification is not achieved without cost. One limitation results directly from the discrete nature of the reservoir size classification employed: the specified categories provide only an approximation to the underlying but unknown continuous distribution. The danger in using any discrete categorization is that some information available in the historical discovery sequence may be lost, because reservoirs of varied but similar sizes are grouped together. The alternative of specifying a continuous form for this distribution (e.g., lognormal) has the advantage of using all of the information available, but even then it can be used only in conjunction with a distributional form that is specified exogenously.

However, the discretized scheme may have a compensating advantage. Reservoir sizes reported in the historical data are known to be subject to considerable error and approximation. Consequently, there is in fact less information available than is apparent. Any classification scheme that is too finely calibrated (including continuous forms) stands the danger of registering many distinctions in reservoir size that are illusory. In other statistical contexts the grouping of similar observations is known to be an effective estimation strategy for controlling errors of measurement in the data. In this case also it may be reasonable to expect that errors of mis-classification are mitigated by using fairly broad size categories. However, the question of what an "optimal" grouping would look like has not been addressed.

 $-14-$ 

On balance, the method of analysis presented here appears to be an important and useful. addition to the set of tools available for economic studies of petroleum supply. It would be useful to further explore the sensitivity of results to the particular form of the postulated sampling mechanism. The concept of discovery without replacement and proportional to size is but one possible characterization of the explorationist's search behavior. There may be other characterizations with even greater intuitive appeal or empirical validity. While the present modelling approach is flexible enough to incorporate alternative specifications of the discovery process, no alternative formulations have yet been explored.

 $\mathcal{L}$ 

-15-

### Appendix: The Data

Estimates of recoverable reserves of oil and gas equivalent (Table 4) were taken from the recent study by Nehring [15] whenever such data were available. Estimates for remaining fields were derived by giving equal proportionate weight to the data reported by Wood, Mackenzie & Co. [21], Riggs National Bank [19], and Robinson and Morgan [20], whenever they were available. If none were available, the estimates reported by John S. Herold, Inc. [9 ] were used. If that were not available, Beall's  $[3]$  estimates were used.

Some discoveries have been reported for which none of the six sources provide an estimate of recoverable reserves. The reserves in such cases were assumed to be small -- less than 50 million barrels. A more detailed description of the reserve estimates is available upon request from the author.

The date of discovery of each reservoir was taken from [21] whenever available. Otherwise, the discovery dates reported in [19] were used. The discovery sequence departs only once from the chronological record, that relating to the Statfjord field. Although Statfjord was not discovered until February, 1974; it is widely believed that the industry had desired to drill the structure much earlier, at approximately the same time as the Brent field, and was only prevented from doing so by the reluctance of the Norwegian government. Therefore, it seems appropriate to place the Statfjord discovery immediately after Brent in the discovery sequence. The effect of this resequencing is small.

#### $-16-$

# TABLE 4: CHRONOLOGICAL DISCOVERY SEQUENCE



### References

- 1. Barouch, E. and G. M. Kaufman, "Oil and Gas Discovery Modelled as Sampling Proportional to Random Size," M.I.T. Sloan School of Management, Working Paper WP888-76, December, 1976.
- 2. Barouch, E. and G. M. Kaufman, "Probabilistic Modelling of Oil and Gas Discovery," in Energy: Mathematics and Models, SIAM Institute for Mathematics and Society, 1976.
- 3. Beall, A. O., "Dynamics of Petroleum Industry Investment in the North Sea," M.I.T. Energy Laboratory Working Paper No. 76-007, June, 1976.
- 4. Eckbo, P. L., H. D. Jacoby, and J. L. Smith, "Oil Supply Forecasting: A Disaggregated Process Approach," Bell Journal of Economics, Vol. 9, No. 1 (Spring, 1978).
- 5. Epple, D., "An Econometric Model of Petroleum Supply with Optimal Endogenous Depletion," paper presented at the Chicago meetings of the Econometric Society, August, 1978.
- 6. Erickson, E. W., and R. M. Spann, "Supply Price in a Regulated Industry: The Case of Natural Gas," Bell Journal of Economics, Vol. 2, No. 1 (Spring, 1971).
- 7. Eyssell, J. H., "The Supply Response of Crude Petroleum: New and Optimistic Results," Business Economics, May, 1978.
- 8. Fisher, F. M., Supply and Cost in the U.S. Petroleum Industry: Two Econometric Studies, Baltimore: Johns Hopkins Press, 1964.
- 9. John S. Herold, Inc., Petroleum Outlook, Vol. 31, No. 8. Greenwich, Connecticut, August, 1978.
- 10. Kaufman, G. M., Y. Balcer, and D. Kruyt, "A Probabilistic Model of Oil and Gas Discovery," in Studies in Petroleum Exploration, No. 1, AIME, 1975.
- 11. Khazzoom, J. D., "The F.P.C. Staff's Econometric Model of Natural Gas Supply in the United States," Bell Journal of Economics, Vol. 2, No. 1 (Spring, 1971).
- 12. MacAvoy, P. W. and R. S. Pindyck, The Economics of the Natural Gas Shortage (1960-1980), Amsterdam: North Holland, 1975.
- 13. Mansvelt-Beck, F. W., and K. M. Wiig, The Economics of Offshore Oil and Gas Supplies, Lexington, Mass.: Lexington Books, 1977.

## References (continued)

- 14. National Petroleum Council, U.S. Energy Outlook, December, 1972.
- 15. Nehring, R., Giant Oil Fields and World Oil Resources, Santa Monica: Rand Corporation, 1978.
- 16. Odell, P. R. and K. F. Rosing, "The North Sea Oil Province: A Simulation Model of Development," Energy Policy, December, 1974.
- 17. Odell, P. R. and K. F. Rosing, Optimal Development of the North Sea's Oil Fields, London: Kogan Page, Ltd., 1976.
- 18. Pindyck, R. S., "The Regulatory Implications of Three Alternative Econometric Supply Models of Natural Gas," Bell Journal of Economics, Vol. 5, No. 2 (Autumn, 1974).
- 19. Riggs National Bank, Status Report on the North Sea Development Drilling Market, Washington, D.C., September, 1977.
- 20. Robinson, C. and J. Morgan, North Sea Oil in the Future, London: The MacMillan Press, Ltd., 1978.
- 21. Wood, Mackenzie& Co., North Sea Report, Section Two, Edinburgh, 1978.